

Integrating Mobility Prediction and Resource Pre-allocation into a Home-Proxy Based Wireless Internet Framework

Jonathan Chan[†], Björn Landfeldt[†], Aruna Seneviratne[†] and Pipat Sookavatana^{†‡}

[†]School of Electrical Engineering and
Telecommunication
The University of New South Wales
Sydney 2052, Australia

[‡]Department of Computer Engineering
Mahanakorn University of Technology
51 Cheum Sampan Rd., Nong Chok
Bangkok 10530, Thailand

[jchan, bjornl, aruna, pipats]@ee.unsw.edu.au

Abstract

This paper provides a novel home-proxy based framework which integrates the mobility management and QoS management techniques that have been adopted in the Internet environment. It then combines a mobility prediction and resource pre-allocation scheme with the above home-proxy based framework to provide an efficient mobility management environment capable of supporting real-time services. The viability and efficiency of the scheme is demonstrated through a simplistic analysis.

1. Introduction

The natural evolution of cellular systems and Internet technologies is towards a wireless Internet, which will provide access not only to voice and short message services, but also to all information services, including real-time communications, from anywhere at anytime. However, the convergence of these technologies to provide ubiquitous access has been hampered by three major factors. Firstly, the traditional Internet protocol suite could not support mobility. Secondly, the bandwidth of the wireless access networks available so far has been insufficient. Finally, the best effort nature of the Internet services, together with the volatility of the radio propagation characteristics have made it impossible for these systems to provide the quality of service (QoS) guarantees necessary to support any real-time services.

Nevertheless, the research community has addressed some of these issues and the techniques are being standardised. The IETF has adopted Mobile IP [1] to handle mobility in Internet environments. There has also been tremendous advances in radio technologies, which has enabled mobile systems to offer data services at much higher data rates in the next generation systems [2]. Moreover, there has been considerable progress in developing techniques for providing QoS guarantees in Internet environments [3, 4]. But these developments,

until recently, have been largely done in isolation. As a result, the mobility management and QoS management techniques that have been adopted are incompatible, and do not provide an integrated solution to the problem of providing real-time services.

The objective of this paper is to describe how a home-proxy based framework together with user mobility prediction and resource pre-allocation can be used to provide an integrated solution to the above problems. Unlike many proposals in this area, our work focuses on the practical perspectives of user movement estimation and resource pre-allocation. The rest of the paper is organised as follows. In Section 2, we give an overview of related work in integrating user mobility into QoS management. In Section 3, we describe the findings of some realistic movement traces and a practical solution to improve the prediction accuracy ratio. In Section 4, we apply this approach to pre-allocate resources in our home-proxy based wireless Internet framework. In Section 5, we present a simplistic analysis that estimates its improvement in handover latency. Finally, we summarise this article in Section 6.

2. Related Work in Incorporating User mobility into Internet QoS Management

It is generally difficult to promise a specified level of QoS to a mobile user since there may not be enough resources in the part of the network that the mobile user is moving into. Even with the assurance of spared network resources, the latency associated with resource allocation at new paths can be prohibitively large, especially when the corresponding host is some distance away. For continuous services in real-time, mobile systems need to provide seamless roaming capability to mobile users. Moreover, the mobile service models should also honour, at least statistically, the QoS agreements between a mobile host and the serving network. The approaches taken to this problem are based on advance reservation. Advance reservation needs to deal with two aspects,

namely how to configure resources in advance, and where to reserve resources for mobile users.

2.1. How to Configure Resources in Advance

Recently, the research community has addressed this issue using a combination of Mobile IP and IntServ models. Depending on the topology used for reserved data paths, these proposals can be classified into three categories: pre-configured anchor rerouting (e.g. MRSVP [5]); pre-configured path extension (e.g. advanced reservation signalling [6]); and pre-configured tunnelling tree (e.g. simple QoS signalling [7]). In many of these approaches, new concepts such as RSVP proxy agents and passive reservation are introduced. Unfortunately, this implies that all intermediate RSVP-capable routers have to be upgraded and it is unclear how these features would impact the IntServ and DiffServ architectures.

2.2. Where to Pre-allocate Network Resources

It is non-trivial to determine where to pre-allocate resources because the future times and locations of a mobile user is very hard to predict. Despite the difficulties, many resource pre-allocation algorithms have been proposed to safeguard the QoS agreements of mobile services. Locations participating in the pre-allocation process can be derived from the surrounding neighbours of the current position [6], the movement history of mobile users [5], or an aggregate virtual system [8]. It is noticeable that the further ahead a scheme tries to pre-allocate resources, the more likely a network can honour an agreement for the lifetime of the session. However, this is achieved at the expense of poor overall network utilisation because of the poor prediction accuracy.

Future mobile service models need to make a compromise between the level of service commitments and the risk of over-subscriptions in the mobile network. The aggregate virtual system appears to be a scalable approach, but the scope of this system and the feasibility of its aggregated functions are unclear [8]. The neighbourhood allocation scheme is the simplest. However, resources are likely to be over-subscribed because users seldom walk randomly in real-life. The allocation scheme based on user mobility history, on the other hand, can reserves resources in selective surrounding cells, and thereby attempts to minimise the

probability of over-subscription in the mobile network. Although this view has been supported by simulation results from various sources [9, 10], its usefulness in reality cannot be fully verified until the actual user mobility is better understood.

3. Mobility Prediction: A Realistic Study

From the above discussion, the resource allocation scheme based on user mobility history seems to be a promising approach. However, most algorithms proposed so far have used simplified movement models that do not accurately characterise user mobility. As a result, these rough assumptions lead to unrealistic conclusions that tend to misplace network resources. Therefore, we believe that it is beneficial to investigate some actual mobility traces and compare our findings with some common movement models.

A commonly perceived notion is that user movements have some level of regularity. In particular, it is believed that by logging the user movements for a sufficiently long time, one can identify the similarities between the current trajectory and the past movement patterns, so that the future location(s) of a mobile user can be determined with a higher confidence.

3.1. Traces of Realistic Movements

In order to verify the above claim, we analysed two real-life user mobility traces in both indoor [11] and outdoor [12] environment. This paper consolidates our previous work by examining another set of outdoor movement traces taken from a local GSM network. In April 1999, we logged the identity of basestations a mobile phone was connected to as a user was travelling by train between the central business district of Sydney and one of its outer suburbs. This user followed the same railway link for both inbound and outbound journeys, and the train regularly stopped at 14 train stations as illustrated in Figure 1. Also, this diagram shows some triangular symbols that represent the locations of basestation registered in the traces. We had traced 26 round-trip journeys during the office hours of five consecutive working days and on average, 30 handover events were recorded for each single trip.

3.2. Prediction of User Movements

Recently we analysed and compared these three sets of mobility traces. Our findings consistently indicate that user



Figure 1. The GSM Movement Traces

mobility patterns displayed a significant amount of variation, and they could not be accurately predicted. In addition, it is noticeable that the matching criteria and the source of mobility history affected the accuracy of a prediction algorithm. In particular, we found that more complicated matching algorithms do not guarantee an improvement on the prediction accuracy, but a simple algorithm based on the direction of travel has the best performance (approximately 72%).

In the indoor movement case, the daily activities of multiple users were captured. Our analysis supports the common assumption that each user has some level of regularity in their movements. In both outdoor movement traces, a single user was travelling forwards and backwards along a fixed route. Despite his regular behaviour, we found that the mobility patterns of these trips varied significantly and they were difficult to predict. By examining these mobility patterns in details, it was found that most “random components” of the traces came from the Ping-Ponging between adjacent base stations or some temporary handovers to other base stations relatively far apart (i.e. not neighbouring base stations). We can visualise this observation by dividing the route in Figure 1 into six segments, and surrounding those basestations at each segment with an ellipse (identified as region A to F in the diagram). For instance, it can be shown that when the user was travelling in region A, his mobile phone tended to handover to a lot more basestations than it did in other regions. We believe that the above effects are caused as a result of a combination of signal fluctuations, constraints of the surroundings, congested cells, and moving obstacles.

Another common assumption of these outdoor mobility patterns is that the forward and backward movement traces are very similar except their order is reversed. In spite of the presence of some key basestations along the path, our analysis does not strongly support this claim of reverse mapping between forward and backward movements. We believe that this variance is a consequence of the use of hysteresis in handover algorithm. In its presence, handover decisions of forward and backward journey are very likely to take place at different locations within a cell, and thereby the statistics of handover patterns tend to vary.

From the above discussion, we have seen a large discrepancy between the user mobility models in the literature and the actual system measurements. The main reason of this problem is that frequent assumptions of mobility model focus on the behavioural patterns (i.e. the physical movements) of users, but pay less attention to user mobility from the perspective of a wireless network. What is required in a wireless system is for the mobility prediction algorithms to predict the network access point through which the mobile user will connect to the network. In other words, we should predict the

cell/basestation to which the user will next connect, rather than the physical location that the user is moving into.

3.3. A Practical Resource Reservation Scheme Based on Mobility History

From our previous study of movement traces, it is clear that a useful mobility prediction algorithm needs to incorporate not only user behavioural patterns, but also wireless link characteristics and the handover decision-making mechanism. However, all these are complicated issues, and comprehensive solutions may not be available in the near future.

Nevertheless, we believe that a user’s movement history still contains valuable information about his/her behavioural patterns and some useful hints about the stability of wireless link along the movement path. As a practical approach to improve the prediction accuracy and the feasibility of resource reservation, we have proposed an adaptive mobility prediction algorithm [11, 12]. Because of the uncertainty of estimating user’s future movements in reality, we have constrained this algorithm to predict only the next move of the user’s present path. A parameter called the Prediction Confidence Ratio (PCR) is used to express the user’s desire of service continuity. This prediction algorithm is defined as follow [11]:

“A prediction is derived from the probability distribution of all possible next moves from the mobility history. If the first predicted cell does not contain a probability higher than the PCR, one or more extra cells will be added to the group of cells in which resources will be reserved in advance. This process will continue until the sum of their probabilities exceeds the targeted PCR.”

The motivation of this adaptive mobility prediction algorithm is to avoid predicting the random components of user movements. Instead of selecting the most likely cell but not meeting the user’s expectation of service continuity, this algorithm can adaptively reserve resources in advance at multiple nearby cells according to the user behavioural patterns and wireless link characteristics. Applying this algorithm with different values of PCR into our mobility traces, we can improve the prediction accuracy ratio accordingly. For example, the prediction accuracy ratio of the outdoor movement traces is initially 72%. By requesting a PCR of 75%, the prediction accuracy ratio increases to 80% but an average cell participation of 1.6 is required [12].

4. Integration of Resource Pre-allocation into a Wireless Internet Framework

In this section we try to implement the concept of PCR into a home-proxy based wireless Internet framework such that network resources can be allocated in advance at multiple locations. Figure 2 shows the environment in

which this framework is to operate. In this diagram, a mobile host (MH) is communicating with a corresponding host (CH) while it is on the move across various administrative domains. The architecture of this system [13] can be broken down into three components: Home Register (HR), Home Proxy (HP), and Wireless Internet Gateway (WIG). These modules collaborate with each other to fulfil the basic functions of roaming services. Because the resource reservation is closely related to the mobility management in our framework, we will briefly describe its operation in the next section.

4.1. Mobility management of the Framework

To accommodate frequent handovers, our mobility management is divided into two levels:

1. Mobility management between administrative domains (*Macro-Mobility*) is handled at the session layer via a session layer mobility management (SLM) module [14]. For instance, When a connection is made to a MH that is away from home, the HP accepts the connection on behalf of the MH. Then the HP makes a separate connection, via its own SLM module, to the SLM module at the MH. Similar to its peer at the HP, the SLM module at the MH accepts the connection on behalf of the MH, and makes a separate connection to the waiting application. When this MH moves to another domain, the inter-SLM connection is simply torn down and re-established between the HP and MH. As the connection state is retained by SLM and the original connections are shielded from the details of the network, the movement of the MH is transparent to the applications. It is note worthy that the maintenance of inter-SLM connection requires no special mobility support at the underlying QoS architecture.
2. Mobility management within an administrative domain (*Micro-Mobility*) is managed at the network layer. A WIG is placed at the boundary of an access network which serves as the mobility interchange gateway between macro- and micro-mobility. To efficiently route data packets to a MH without tunnelling or address conversion, we have chosen a dynamic per-

host routing scheme similar to HAWAII [15] or Cellular IP [16]. In this approach, packets are forwarded on a hop-by-hop basis from WIG over a dynamically established path to the MH. Since the IP address of MH is fixed, the user mobility within the access network would be transparent to the Internet backbone.

4.2. Resource Pre-allocation of the Framework

To reduce the complexity of resource reservation for MHs, we assume that in this framework the bottleneck of network resources is likely to appear at the Internet backbone or at the wireless link. This implies that the network paths joining BSs and the WIG should be either well engineered or over provisioned. We believe that it is not difficult to achieve in real practice because the internal network is normally owned by the same administrative entity and has relatively large and cheap bandwidth. Based on this assumption, we can divide resource reservation into two levels, similar to the mobility management.

1. *Advance Macro-Reservation*: WIG plays an important role in our resource reservation scheme for mobile users. At the macro-level, it behaves as the first or last hop of the underlying QoS framework to complete the QoS path across the Internet backbone. For instance, it can act as an ingress or egress router for the DiffServ architecture, or as a RSVP proxy agent [5, 6] for the IntServ architecture. Advance macro-reservation can be achieved without further changes in the existing QoS framework, because the SLM module at the HP can simply establish multiple connections to different WIGs if their domains contain the predicted locations of next move.
2. *Advance Micro-Reservation*: At the micro-level, the WIG allocates wireless resources at BS(s) in which the QoS path is expected to be initiated or terminated. In our framework, the QoS support across the Internet backbone is managed by the macro-reservation. Therefore, we argue that by assigning wireless resources for a QoS path at multiple BSs, we are effectively extending the QoS support of this Internet connection in advance into multiple locations. Moreover, as the mobile user moves from one location to another, the basestations participating in resource reservation can vary according to the prediction. Nevertheless, regardless of the user location and mobility patterns inside the domain, the same QoS path is used to carry traffic across the Internet backbone, and hence no change is required at macro-reservation level.

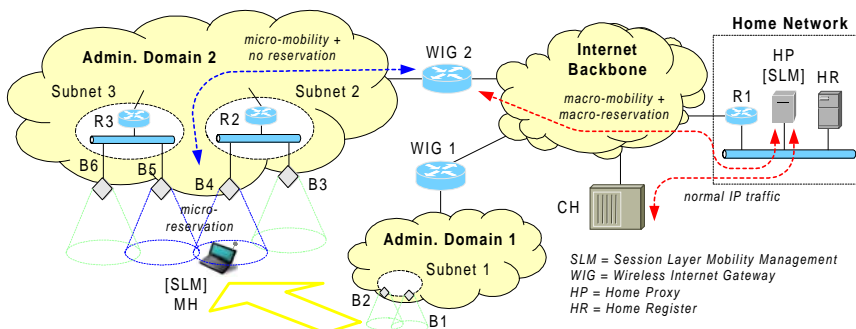


Figure 2. The dual-level approach of mobility management and advance resource reservation in the home-proxy based wireless Internet framework

These different levels of resource pre-allocation processes are schematically shown in Figure 3. For the ease of illustration, we only indicate one possible location of next move in the diagram, but this reservation process can be repeated multiple times depending on the PCR value and the probability distribution of the next move.

5. Improvement of Handover Latency using Advance Reservation

In this section we present a simplistic analysis on the duration of handover latency with and without advance reservation. The process of macro- and micro-mobility management without advance reservation is very similar to the one shown in Figure 3. However, the MH does not send out any pre-allocation request and hence resource reservation and DHCP address assignment will start after the MH has arrived at the new location.

In the analysis, we calculate the duration of various stages during a handover (i.e. Figure 3 (b) – (e)) by adding up the delay of subsequent protocol entities. It is assumed that all control and QoS signalling messages travel at the same speed (2ms locally and 50ms across the backbone), and processing delays are negligible at all

network nodes. Note that the delay of DHCP assignment and the delay of arrival detection depend heavily on the software implementation. It was measured that the DHCP assignment can vary from 2.6 seconds (standard implementation) to 103ms (optimised) [17]. Similarly, the detection of subnet advertisements can range from 2.5 seconds (standard implementation) to a few msec (say 5ms via link layer detection) [18].

The result of our analysis shows that the delay of arrival detection has a negative impact on handover latency for both the macro- and micro-mobility. The handover latency of macro-mobility without advance reservation can be as high as 5.26 seconds, whereas the latency with advance reservation can be reduced to 2.61 seconds. This improvement is caused by the pre-assignment of DHCP address during advance macro-reservation.

Provided if the arrival detection is handled by link layer beacons, the handover latency of macro-mobility can be as small as 113ms for advance macro-reservation. This value is roughly equal to the round trip delay of the Internet backbone. However, in the scenario of macro-mobility without any advance reservation, the handover latency still takes about 2.77 seconds because DHCP

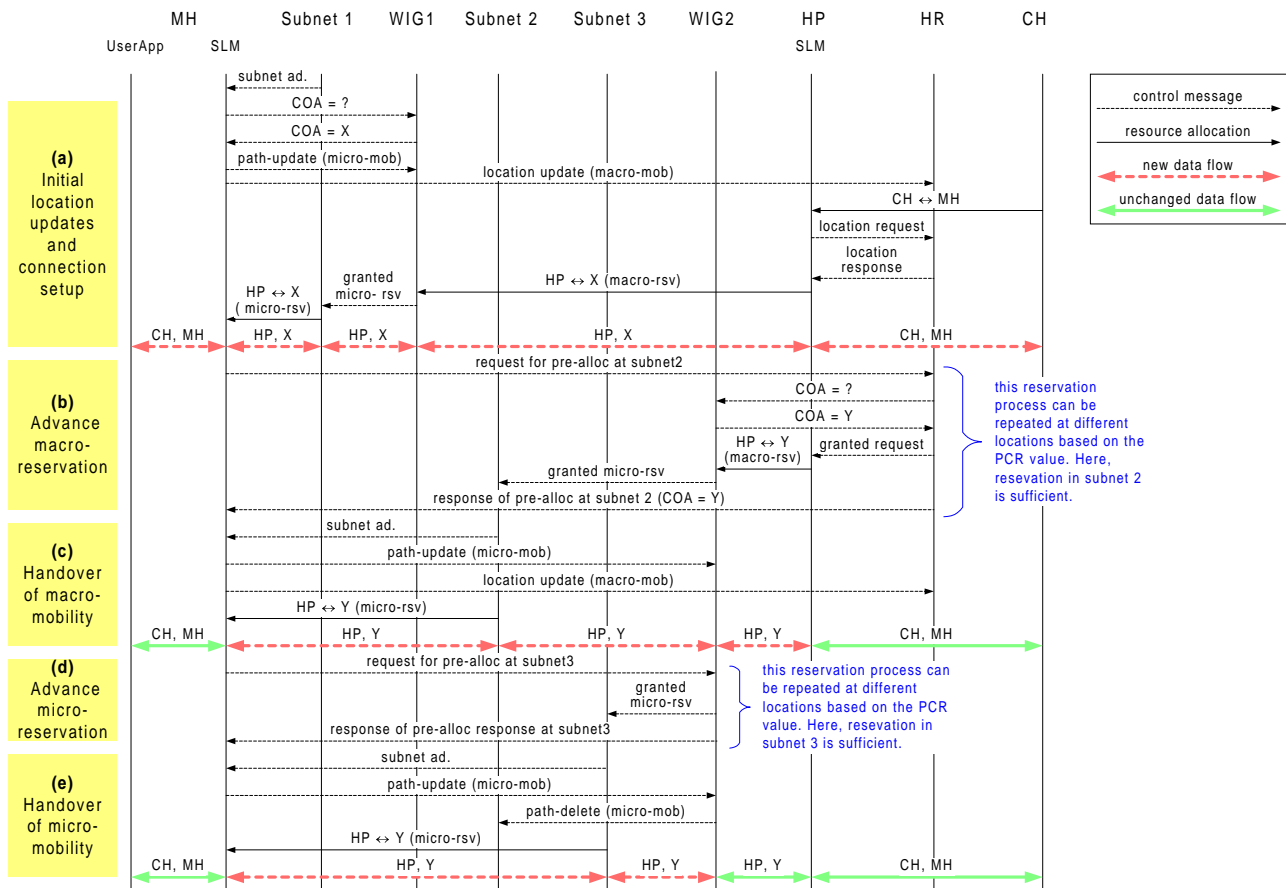


Figure 3. The schematics of advance micro- and macro-reservation using the concept of PCR

address allocation is performed after the arrival.

Regarding the handover latency of micro-mobility, the delay of arrival detection still plays an important role. Once being handled by link layer beacons, it can be as small as 11ms and 17ms for handover latency with and without advance reservation.

From the above calculation, it is noticeable that the delays of arrival detection and DHCP assignment should be optimised if real-time services are supported in the network. Finally, one important aspect not mentioned in this analysis is that with advance reservation, service continuity would be secured after the handover.

6. Summary

In this paper we have presented an overview of our home-proxy based mobility management scheme. Furthermore, we have investigated some user mobility traces. The results show that user mobility patterns in real life are influenced by a combination of signal fluctuations, constraints of the surroundings and handover mechanism. To pre-allocate network resources in reality, we should predict the cell to which the user will next connect, rather than the physical location that the user is moving to. We have proposed the Prediction Confidence Ratio (PCR) through which network resources can be reserved in advance at multiple cells according to the user mobility patterns. Then, we have applied this concept to our home-proxy based framework and formed a dual-level approach of mobility management and resource pre-allocation. Compared with other proposed schemes of advance reservation in Section 2, our techniques appear to be more feasible with little complication to the underlying QoS architecture. Finally, through a simplistic analysis we have identified the importance of optimising the delay of arrival detection, and the advantages of performing advance reservation during handover. To prove the viability of the framework as mentioned in this paper, an experimental testbed is currently being built.

Acknowledgements

The work of Jonathan Chan is funded through a CSIRO Postgraduate Scholarship and an Australian Postgraduate Award. The work of Björn Landfeldt is funded by Ericsson, Australia.

References

- [1] C. Perkins, "IP Mobility Support for IPv4, revised" *IETF Internet Draft*, Oct. 1999.
- [2] T. Ojanperä and R. Prasad, "An Overview of Third-Generation Wireless Personal Communications: A European Perspective," *IEEE Pers. Commun. Mag.*, Dec. 1998.
- [3] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," *IETF RFC 1633*, Jun. 1994.
- [4] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differential Services," *IETF RFC 2475*, Dec. 1998.
- [5] A. Talukdar, B. Badrinath, and A. Acharya, "MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts," to appear in *ACM/WINET*, May 1999.
- [6] I. Mahadevan and K. Sivalingam, "Architecture and Experimental Results for Quality of Service in Mobile Networks using RSVP and CBQ," to appear in *ACM/WINET*, Jul. 1999.
- [7] A. Terzis, M. Srivastava, and L. Zhang, "A Simple QoS Signaling Protocol for Mobile Hosts in the Integrated Services Internet," in *Proc. IEEE INFOCOM'99*, Mar. 1999.
- [8] R. Jain, B. Sadeghi, and E. Knightly, "Towards Coarse-Gained Mobile QoS," in *Proc. ACM WoWMoM'99*, Aug. 1999.
- [9] C. Oliveira, J. Kim, and T. Suda, "An Adaptive Bandwidth Reservation Scheme for High-Speed Multimedia Wireless Networks," *IEEE JSAC*, 16(6), Aug. 1998.
- [10] P. Ramanathan, K. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless Networks," *IEEE JSAC*, 17(7), Jul. 1999.
- [11] J. Chan, S. Zhou and A. Seneviratne, "A QoS Adaptive Mobility Prediction Scheme for Wireless Networks," in *Proc. IEEE GLOBECOM'98*, Nov. 1998.
- [12] J. Chan and A. Seneviratne, "A Practical User Mobility Prediction Algorithm for Supporting Adaptive QoS in Wireless Networks," in *Proc. IEEE ICON'99*, Sep. 1999.
- [13] J. Chan, B. Landfeldt, R. Liu, and A. Seneviratne, "A Home-Proxy Based Wireless Internet Framework in Supporting Mobility and Roaming of Real-Time Services," submitted for review, Jul. 2000.
- [14] B. Landfeldt, T. Larsson, Y. Ismailov, and A. Seneviratne, "SLM, A Framework for Session Layer Mobility Management," in *Proc. IEEE ICCCN*, Oct. 1999.
- [15] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, and L. Salgarelli, "IP Micro-mobility Support using HAWAII," *IETF Internet Draft*, Jun. 1999.
- [16] A. Valkó, "Cellular IP: A New Approach to Internet Host Mobility," *ACM Comp. Commun. Review*, Jan. 1999.
- [17] J-O Vatn and G. Maguire Jr., "The effect of using co-located care-of addresses in macro handover latency," Fourteenth Nordic Tele-traffic Seminar (NTS 14), Aug. 1998.
- [18] A. Fladenmuller and R. De Silva, "The effect of Mobile IP handoffs on the performance of TCP," *ACM/Baltzer MONET*, 4(2), 1999.